

# Refining a Quantitative Information Flow Metric

Sari Haj Hussein

Department of Computer Science, Aalborg University, Denmark  
Email: angyjoe@gmail.com

**Abstract**—We introduce a new perspective into the field of quantitative information flow (QIF) analysis that invites the community to bound the leakage, reported by QIF quantifiers, by a range consistent with the size of a program’s secret input instead of by a mathematically sound (but counter-intuitive) upper bound of that leakage. To substantiate our position, we present a refinement of a recent QIF metric that appears in the literature. Our refinement is based on slight changes we bring into the design of that metric. These changes do not affect the theoretical premises onto which the original metric is laid. However, they enable the natural association between flow results and the exhaustive search effort needed to uncover a program’s secret information (or the residual secret part of that information) to be clearly established. The refinement we discuss in this paper validates our perspective and demonstrates its importance in the future design of QIF quantifiers.

**Index Terms**—computer security, quantitative information flow, information theory, uncertainty, inference, program analysis

## I. INTRODUCTION

The goal of information flow analysis is to enforce limits on the use of information that apply to all computations that involve that information. For instance, a confidentiality property requires that a program with secret inputs should not leak those inputs into its public outputs. Qualitative information flow properties, such as non-interference are expensive, impossible, or rarely satisfied by real programs; generally some flow exists, and many systems remain secure provided that the amount of flow is sufficiently small, moreover, designers wish to distinguish acceptable from unacceptable flows.

Systems often reveal a summary of secret information they store. The summary contains fewer bits and provides a limit on the attacker’s inference. For instance, a patient’s report is released with the disease name covered by a black rectangle. However, it is not easy to precisely determine how much information exists in the summary. For instance, if the font size is uniform on the patient’s report, the width of the black rectangle might determine the length of the disease name. Quantitative information flow (QIF) analysis is an approach that establishes bounds on information that is leaked by a program. In QIF, confidentiality properties are also expressed, but as limits on the number of bits that might be revealed from a program’s execution. A violation is declared if the number of leaked bits exceeds the policy. Because information theory forms the foundation of QIF analysis, it should be possible to associate the quantities reported by QIF quantifiers with the effort needed to uncover secret information via exhaustive search. However, establishing this association is infeasible with QIF quantifiers that do not report a flow *consistent* with the

size of a program’s secret input, but instead a mathematically sound *upper bound* of that flow [1]. For instance, consider the QIF metric and the password checker in Section 1 of [1], and assume that the password space has a cardinality of 3. This means that the size of the password is  $\log 3 = 1.5849$  bits. (Here and hereafter, all logarithms are to the base 2). Nonetheless, the metric in [1] might report a flow that exceeds 1.5849 bits, which makes it impossible to determine the space of the exhaustive search that should be carried out in order to reveal the residual secret part of the password. However, if the flow reported is always less than 1.5849 bits, the exhaustive search space becomes evident.

We believe that the counter-intuitive flow quantities reported by some QIF quantifiers, that appear in the literature, are due to a flaw in the design of those quantifiers, and that simple tweaks can bound those quantities by a range consistent with the size of a program’s secret input. This paper takes the first step in this direction and refines the QIF metric suggested in [1]. The metric in [1] is based on a new perspective for QIF analysis. The fundamental idea is to model an attacker’s belief about a program’s secret input as a probability distribution over high states. This belief is then revised, using Bayesian updating techniques, as the attacker interacts with a program’s execution. It is believed that the work reported in [1] is the first to address an attacker’s belief in quantifying information flow. This work was later expanded and appeared in [2]. A number of relevant results [3], [4] were reported in the sequel; however, the work in [1], [2] is sufficient as a foundation of our work.

### A. Plan of the Paper

The remainder of this paper is organized as follows. Section II elaborates on accuracy-based information flow analysis which is the major contribution in [1]. In this section, we give concise elucidation of the elements of this analysis and how it differs from the classical uncertainty-based information flow analysis. In addition, we uncover some inexplicable results reported by the QIF metric in [1], and argue that the reasoning of this metric’s designers is *incomplete*. We further state the general range of flow reported by the metric in [1] that applies to both deterministic and probabilistic programs as well as to all types of attacker’s beliefs. This range is neither given in [1] nor in [2]. Over the course of acquiring the range, we reveal the ineffectiveness of the admissibility restriction suggested in [1]. At the end of Section II, we conjecture a simple fix that can bound the results reported by the metric in [1]. Underpinning our arguments in Section II is a formal definition of a *size-consistent QIF quantifier*. Our definition

is based on uncertainty-based information flow analysis, and it inaugurates the new perspective we are introducing into the field of QIF. To the best of our knowledge, this is the *first* definition to capture the correlation between the size of a program's secret input and the quantification of flow from that input in the general case. Section III concentrates on Kullback-Leibler divergence which is a centerpiece of the metric in [1]. We give some mathematical interpretations of this divergence, and then focus on its discrimination construct, suggesting the replacement of this construct with a better one, and subsequently the replacement of the divergence itself with another, bounded, divergence. This paves the way for the refinement of the metric in [1] which is what we fulfill in two stages in Section IV. We also give the range and the interpretation of the refined metric, and prove its properties and their meaningfulness compared to the original one, while minding the consistency of the probability distributions dealt with. Having justified the conjecture we made in Section II, and shown that a large number of possible refinements of the metric in [1] exist, we discuss the association of the original and the refined metric with the exhaustive search effort in Section V, give some remarks in Section VI, and conclude the paper in Section VII. The proofs are given in Appendix I.

## II. UNCERTAINTY- VS. ACCURACY-BASED INFORMATION FLOW ANALYSIS

The problem with uncertainty-based information flow analysis is that it *ignores* reality. As an example, consider a simple password checker  $\mathcal{PWC}$  [1] that sets an authentication flag  $a$  after checking a stored password  $p$  against a guessed password  $g$  supplied by the user.

$$\mathcal{PWC} : \text{ if } p = g \text{ then } a := 1 \text{ else } a := 0 \quad (1)$$

For simplicity, suppose that the password space is  $\mathcal{W}_p = \{A, B, C\}$ , which gives a size of  $\log |\mathcal{W}_p| = \log 3 = 1.5849$  bits for the password  $p$ . Suppose further that the user is actually an attacker attempting to discover the password. *Before* interacting with a  $\mathcal{PWC}$  execution, this attacker believes that the password is *overwhelmingly* likely to be  $A$  but has a very small and equally likely chance to be either  $B$  or  $C$ . More concretely and adopting the convention in [1], the attacker's *prebelief* about  $p$  is captured using a probability distribution  $b_H : \mathcal{W}_p \rightarrow [0, 1]$  as shown in Table Ia.

$p$	$A$	$B$	$C$
$b_H$	0.98	0.01	0.01

(a) Attacker's prebelief

$p$	$A$	$B$	$C$
$b'_H$	0	0.5	0.5

(b) Attacker's postbelief

TABLE I: Attacker's beliefs in the password  $p$

The attacker's uncertainty about  $p$  (*not* necessarily about the correct  $p$ ) is obtained via a simple application of Shannon uncertainty functional [5]:

$$\mathcal{U} = S(b_H) = -0.98 \log 0.98 - 2 \cdot 0.01 \log 0.01 = 0.1614 \text{ bits}$$

Assuming that the correct password (the *reality*) is  $C$ , if the attacker complies to her prebelief and feeds a  $\mathcal{PWC}$  execution with  $g = A$ , *she* will observe  $a$  equal to 0. The attacker then infers that  $A$  is not the real password, and that there is an equal chance of 50% that the password is either  $B$  or  $C$ . As a result, the attacker's postbelief distributes as shown in Table Ib, and the attacker's uncertainty about  $p$  becomes:

$$\mathcal{U} = S(b'_H) = -0.5 \log 0.5 - 0.5 \log 0.5 = 1 \text{ bit}$$

To complete an uncertainty-based information flow analysis, we have to compute the *reduction* in uncertainty by subtracting the post- from the pre-uncertainty using the formula:

$$\mathcal{R} = \mathcal{U} - \mathcal{U}'$$

This gives us  $\mathcal{R} = 0.1614 - 1 = -0.8386$  bits. In the sense of uncertainty-based analysis, the negative  $\mathcal{R}$  means *absence* of information flow. There is *nothing* wrong with this interpretation provided that we do not connect information flow with how *far* an attacker's belief is from reality. However, if we connect the flow with the distance between an attacker's belief and reality, then the interpretation that  $\mathcal{R}$  supports does not make sense. The measure  $\mathcal{R}$  ignores reality by measuring  $b_H$  and  $b'_H$  against each other only, instead of against the high state (which is  $C$  as the correct password in our example). It is good to notice however that the range of flow reported by  $\mathcal{R}$  is as given by the formula:

$$\varrho \mathcal{R} = [-\log |\mathcal{W}_p|, \log |\mathcal{W}_p|] = [-1.5849, 1.5849]$$

This is a direct consequence of Shannon uncertainty functional falling in the range  $[0, \log |\mathcal{W}_p|]$  [6]. The range  $\varrho \mathcal{R}$  reported by  $\mathcal{R}$  is *plausible* if we remember that the size of the password  $p$  is 1.5849 bits. We would like to take time defining the size-consistent QIF quantifier.

*Definition 1 (Size-consistent QIF Quantifier):* We say that a QIF quantifier is size-consistent if its reported results are bounded (from above and from below) by the size of a program's secret input. Formally, let  $\mathcal{QUAN}$  be a QIF quantifier, and assume that the size of a program's secret input is  $\eta$  bits. We say that  $\mathcal{QUAN}$  is size-consistent if:

$$\mathcal{QUAN}_{max} \leq \eta \text{ and } \mathcal{QUAN}_{min} \geq -\eta$$

However, if we merely look at the attacker's prebelief and postbelief in  $C$ , as the correct password, we realize that the attacker's belief has *approached* reality from interacting with  $\mathcal{PWC}$ . Approaching reality cannot happen unless the attacker *learns* something from an amount of information  $\mathcal{PWC}$  has conveyed. This conveyance corresponds to *positive* information flow that informs the attacker, and flatly contradicts the uncertainty-based interpretation.

The earliest investigation of this *specific* inadequacy of uncertainty-based information flow analysis appeared in [1] and was later expanded in [2]. The authors of [2] propose to respect reality through what they call "accuracy-based information flow analysis". This sort of analysis has two elements:

- E1. Quantifying information flow from a program's execution to an attacker.

E2. Respecting the distance between an attacker's belief and reality.

The uncertainty-based analysis does *not* have the second element as the example above demonstrated. The accuracy-based analysis quantifies flow as the *improvement* in the accuracy of an attacker's belief. This is equivalent to saying the reduction in the distance between an attacker's belief and reality. The metric advanced in [2] is based on this notion of improvement, and is given by the formula:

$$\mathcal{Q}(\mathcal{E}, b'_H) = D(b_H \rightarrow \dot{\sigma}_H) - D(b'_H \rightarrow \dot{\sigma}_H) \quad (2)$$

where  $\mathcal{E} = \langle S, b_H, \sigma_H, \sigma_L \rangle$  is an experiment tuple as defined in [2],  $\langle \mathcal{E}, b'_H \rangle$  is the outcome of that experiment,  $b_H$  is the attacker's prebelief,  $b'_H$  is the attacker's postbelief,  $\dot{\sigma}_H$  is a probability distribution that maps the high state  $\sigma_H$  to 1 (this is the *certainty* about the high state; about *reality*), and  $D$  is Kullback-Leibler divergence (also known as relative entropy or information gain [6]) given by the formula:

$$D(b \rightarrow b') = \sum_{\sigma \in \mathcal{W}_p} b'(\sigma) \cdot \log \frac{b'(\sigma)}{b(\sigma)} \quad (3)$$

Notice in formula (2) how  $\mathcal{Q}$  respects reality by measuring  $b_H$  and  $b'_H$  against the correct high state  $\dot{\sigma}_H$ , instead of against each other only. Formula (2) is simplified in [2] to (this simplification is *reality-aware*):

$$\begin{aligned} \mathcal{Q}(\mathcal{E}, b'_H) &= D(b_H \rightarrow \dot{\sigma}_H) - D(b'_H \rightarrow \dot{\sigma}_H) \\ &= \sum_{\sigma \in \mathcal{W}_p} \dot{\sigma}_H(\sigma) \cdot \log \frac{\dot{\sigma}_H(\sigma)}{b_H(\sigma)} \\ &\quad - \sum_{\sigma \in \mathcal{W}_p} \dot{\sigma}_H(\sigma) \cdot \log \frac{\dot{\sigma}_H(\sigma)}{b'_H(\sigma)} \\ &= -\log b_H(\sigma_H) + \log b'_H(\sigma_H) \end{aligned} \quad (4)$$

To complete an accuracy-based information flow analysis parallel to the uncertainty-based analysis we have completed earlier in this section, we apply formula (4) to the same example given above to obtain:

$$\mathcal{Q}(\mathcal{E}, b'_H) = -\log 0.01 + \log 0.5 = 5.6438 \text{ bits} \quad (5)$$

The flow value of 5.6438 bits reported by  $\mathcal{Q}$  violates the plausible range  $\varrho_{\mathcal{R}} = [-1.5849, 1.5849]$  and equally exceeds the size needed to store the password  $p$ . How can a flow from  $p$  exceed the size needed to store  $p$ ? A sound but puzzling result in the *field* of QIF analysis that the authors of [2] attribute to that the attacker's prebelief is not uniform; it is more erroneous than a *uniform* belief ascribing  $1/3$  probability to each password  $A$ ,  $B$ , and  $C$ , and therefore a *larger* amount of information is required to correct it! But what can the source of this larger amount of information be? Is it a *covert* agent external to the system and the attacker when all the agents are assumed condensed to just the attacker and the system [2]? Besides is it always true that a uniform attacker's prebelief would, in a series of experiments, cause her to learn a total of  $\log 3$  bits [2]? This claim is valid for a deterministic password checker, but *incomplete* for a probabilistic one. Let us verify this fact.

It is proved in [2] that for deterministic programs (including the deterministic  $\mathcal{PWC}$  given in formula (1)), we have:

$$b_H(\sigma_H) \leq b'_H(\sigma_H) \quad (6)$$

Since  $b'_H$  is a probability distribution, we can write:

$$b_H(\sigma_H) \leq b'_H(\sigma_H) \leq 1$$

which means:

$$\begin{aligned} \log b_H(\sigma_H) &\leq \log b'_H(\sigma_H) \leq 0 \\ 0 &\leq \mathcal{Q} \leq -\log b_H(\sigma_H) \end{aligned}$$

The attacker's prebelief is assumed uniform on  $\mathcal{W}_p$ , therefore:

$$0 \leq \mathcal{Q} \leq \log 3$$

Thus, it is beyond a shadow of a doubt that a uniform attacker's prebelief would cause her to learn a total of  $\log 3$  bits from interacting with a deterministic  $\mathcal{PWC}$ . But does the attacker's learning outcome differ when interacting with a probabilistic  $\mathcal{PWC}$ ? An illustrative probabilistic  $\mathcal{PWC}$  is:

$$\begin{aligned} \mathcal{PPWC} : \text{ if } p = g \text{ then } a &:= 1_{0.99} \parallel a := 0 \\ \text{else } a &:= 0_{0.99} \parallel a := 1 \end{aligned}$$

The inequality in formula (6) no longer holds, and we are free to write:

$$\begin{aligned} 0 &\leq b'_H(\sigma_H) \leq 1 \\ -\infty &\leq -\log b_H(\sigma_H) + \log b'_H(\sigma_H) \leq -\log b_H(\sigma_H) \\ -\infty &\leq \mathcal{Q} \leq 0 \text{ or } 0 \leq \mathcal{Q} \leq \log 3 \end{aligned}$$

The sub-range  $-\infty \leq \mathcal{Q} \leq 0$  shows that a uniform attacker's prebelief might cause her to learn an infinite number of *misinforming* bits from interacting with  $\mathcal{PPWC}$ . This demonstrates the *incompleteness* of the claim "a uniform attacker's prebelief would, in a series of experiments, cause her to learn a total of  $\log 3$  bits" made in [2].

The previous discussion motivates the investigation of the general range of the  $\mathcal{Q}$  metric that holds with both deterministic and probabilistic programs as well as with all types of attacker's beliefs. This range is attained in Lemma 1.

*Lemma 1:* Considering both deterministic and probabilistic programs, and all types of an attacker's beliefs, the general range of flow reported by  $\mathcal{Q}$  is:

$$\varrho_{\mathcal{Q}} = (-\infty, -\log b_H(\sigma_H)]$$

Clearly  $\mathcal{Q}$  is not size-consistent. Let us now muse on the computation in formula (5) and try to figure out a mean to proceed with this correspondence. The flow of 5.6438 bits has brought the attacker from  $-\log 0.01 = 6.6438$  bits away from reality to  $-\log 0.5 = 1$  bits away from it. In addition and as proved in Theorem 3 in [2], each bit of flow has made the attacker twice as likely to guess *correctly* [7], or equivalently twice as certain about the *correct* high state (in total, we have  $2^{5.6438} \approx 50$  times increase in the likelihood of a correct guess). In the uncertainty-based definition, the attacker's certainty is ascribed to a high state that *might* be incorrect...Conjecture 1 engrossedly stops the correspondence.

*Conjecture 1:* Considering Theorem 3 in [2], if a bit of flow makes the attacker *more* than twice as likely to guess correctly, then  $\mathcal{Q}$  should become size-consistent.

Seeking a justification for this conjecture will be the purpose of the later sections. Although the authors of [1], [2] are acclaimed for their contribution to the field of QIF through their accuracy-based analysis, their metric allows the respect for reality (element E2) to attenuate the quality of flow quantification (element E1). This attenuation is the result of *severe* discrimination in Kullback-Leibler divergence as we shall see in the next section.

### III. CONCENTRATING ON KULLBACK-LEIBLER DIVERGENCE

#### A. Possible Interpretations of the Divergence

The divergence  $D$  between  $b$  and  $b'$ , given in formula (3), can be interpreted in terms of code *inefficiency* as follows;  $D$  is the average number of bits that are wasted by encoding events from a distribution  $b'$  with a code based on a not-quite-right distribution  $b$  [8]. Another way of writing  $D$  in terms of the *expected* value function [9] is as follows:

$$D(b \rightarrow b') = E_{b'}(\log \frac{b'(\sigma)}{b(\sigma)}), \quad E_{b'}(f) = \sum_{\sigma \in \mathcal{W}_p} b'(\sigma) \cdot f(\sigma)$$

The function  $E_{b'}$  takes the weighted average of the values  $f(\sigma)$  in which the weights are probabilities  $b'$ . In the original paper by Kullback and Leibler [10], the values:

$$\mathcal{I}_{Dis}(\sigma) = \log \frac{b'(\sigma)}{b(\sigma)} \quad (7)$$

are seen as the information in  $\sigma$  for the *discrimination* between  $b$  and  $b'$ . This is plausible if we rewrite the previous values as:

$$-\log b(\sigma) - (-\log b'(\sigma))$$

and recall that the information contained in an observation of an event  $E$  with probability  $p(E)$  is  $-\log p(E)$  [6].

This notion of discrimination leads to another interpretation of  $D$ ; it is the weighted average of the information in  $\sigma$  for the discrimination between  $b$  and  $b'$  where the weights are probabilities  $b'$ . We write:

$$D(b \rightarrow b') = E_{b'}(\mathcal{I}_{Dis}(\sigma)) \quad (8)$$

#### B. A Better Discrimination Construct

We propose to replace the discrimination construct in formula (8) with the following:

$$\mathcal{I}'_{Dis}(\sigma) = \log \frac{b'(\sigma)}{\frac{b'(\sigma) + b(\sigma)}{2}} \quad (9)$$

for  $\mathcal{I}'_{Dis}(\sigma)$  to be the information in  $\sigma$  for the discrimination between the mean  $(b' + b)/2$  and  $b'$ . But what is the effect of this replacement? The following lemma shows that we have actually cut down the discrimination *at least* by half.

*Lemma 2:* The proposed discrimination construct cuts down the discrimination in Kullback-Leibler divergence at least by half, that is:  $\mathcal{I}'_{Dis}(\sigma) \leq \frac{1}{2}\mathcal{I}_{Dis}(\sigma)$ .

A graphical comparison between  $\mathcal{I}_{Dis}(\sigma)$  and  $\mathcal{I}'_{Dis}(\sigma)$  is shown in Figure 1a. It is important to notice at this stage that halving the infinite value of  $\mathcal{I}_{Dis}(\sigma)$  does *not* make it finite.

#### C. A Better Divergence

Substituting (9) for (7) in (8), we get the divergence:

$$D'(b \rightarrow b') = \sum_{\sigma \in \mathcal{W}_p} b'(\sigma) \cdot \log \frac{b'(\sigma)}{\frac{b'(\sigma) + b(\sigma)}{2}} \quad (10)$$

The resulted divergence meets with the asymmetric form  $K$  of Jensen-Shannon divergence proposed in [11]. In fact, formula (9) and Lemma 2 both appear in [11] wrapped in the expected value function.  $D'$  is nonnegative and equals zero if and only if  $b = b'$  [11]. This is essential for any measure of difference and *justifies* using  $D'$  instead of  $D$  to measure the distance between two beliefs. A possible interpretation of  $D'$  is as follows; how much information is *lost* if we describe the two random variables that correspond to  $b$  and  $b'$  with their average distribution  $(b' + b)/2$ ? This interpretation gives  $D'$  the nickname "information radius" [8].

A graphical comparison between  $D$  and  $D'$  is shown in Figure 1b. Notice that  $D$  approaches infinity when  $t$  approaches 0 or 1. In contrast,  $D'$  is always well defined in the entire range  $t \in [0, 1]$ . This is because  $(b' + b)/2 \neq 0$  if either  $b' = 0$  or  $b = 0$ . But what is the effect of using  $D'$  instead of  $D$  in  $\mathcal{Q}$ ? This will be our focus in the next section.

### IV. REFINING THE METRIC

#### A. Refining to Normalization

If we substitute (10) for (3) in (2), we get the metric:

$$\begin{aligned} \mathcal{Q}'(\mathcal{E}, b'_H) &= D'(b_H \rightarrow \dot{\sigma}_H) - D'(b'_H \rightarrow \dot{\sigma}_H) \\ &= \sum_{\sigma \in \mathcal{W}_p} \dot{\sigma}_H(\sigma) \cdot \log \frac{\dot{\sigma}_H(\sigma)}{\frac{\dot{\sigma}_H(\sigma) + b_H(\sigma)}{2}} \\ &\quad - \sum_{\sigma \in \mathcal{W}_p} \dot{\sigma}_H(\sigma) \cdot \log \frac{\dot{\sigma}_H(\sigma)}{\frac{\dot{\sigma}_H(\sigma) + b'_H(\sigma)}{2}} \\ &= -\log(1 + b_H(\sigma_H)) + \log(1 + b'_H(\sigma_H)) \end{aligned}$$

Notice that the above substitution does *not* destroy the bedrock of accuracy-based analysis which, as mention in Section II, quantifies flow as the improvement in the accuracy of an attacker's belief. This *guarantees* that  $\mathcal{Q}'$  is a real metric of information flow. Before proceeding any further, we need to investigate the general range of  $\mathcal{Q}'$ , which is what we do in Lemma 3.

*Lemma 3:* Considering both deterministic and probabilistic programs, and all types of an attacker's beliefs, and *avoiding* the imposition of any admissibility restriction on those beliefs, the general range of flow reported by  $\mathcal{Q}'$  is:

$$\rho_{\mathcal{Q}'} = [-1, 1]$$

Fortunately, the sub-range  $[-1, 0]$  corresponds to the attacker's *misinformation* while the sub-range  $[0, 1]$  corresponds to the attacker's *information* about the correct high state.

The new range  $\rho_{\mathcal{Q}'} = [-1, 1]$ , we have reached, does not make  $\mathcal{Q}'$  size-consistent. Nonetheless,  $\rho_{\mathcal{Q}'}$  is a plausible normalization (flow *percentage*) that is invariant with respect to the choice of the measurement unit.

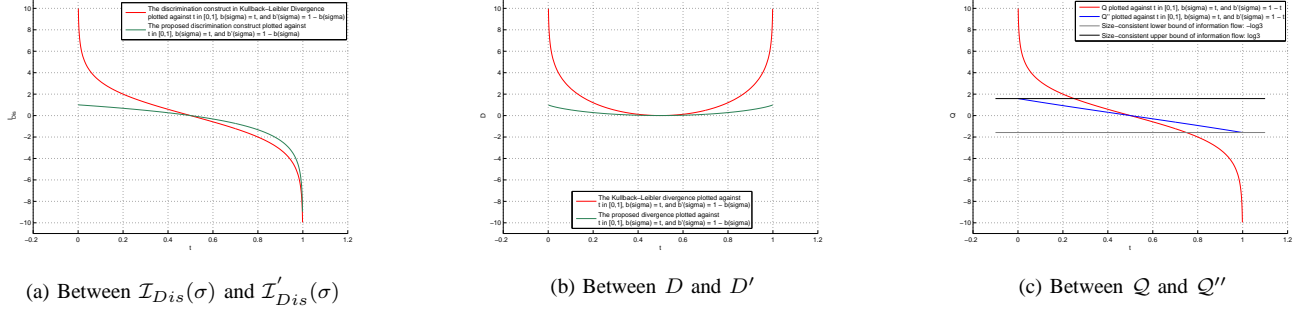


Fig. 1: Graphical comparisons made in the paper

### B. Refining to Actuality

To ensure bits as the measurement unit, and avoid the need to transform the flow results back and forth between the ranges  $\varrho_{Q'} = [-1, 1]$  and  $\varrho_{\mathcal{R}} = [-1.5849, 1.5849]$ , we let  $\eta$  be the size of a program's secret input in bits, and define the refined metric as:

$$\begin{aligned} Q''(\mathcal{E}, b'_H) &= \eta \cdot Q'(\mathcal{E}, b'_H) \\ &= \eta \cdot [-\log(1 + b_H(\sigma_H)) + \log(1 + b'_H(\sigma_H))] \end{aligned} \quad (11)$$

A graphical comparison between  $Q$  and  $Q''$  in the case of *PWC*, along with the size-consistent uncertainty-based upper and lower bounds of flow, is shown in Figure 1c. It is important to notice in this figure that the parts of the  $Q$  and  $Q''$  graphs that fall *above* the zero mark on the  $Y$  axis represent the attacker's information about the correct high state. In contrast, the attacker's misinformation is represented by the parts that fall *below* the zero mark on the  $Y$  axis. Another important observation to make in this figure is that, akin to  $Q$ ,  $Q''$  is *sensitive* to changes in the attacker's belief. It is thus noted that  $Q''$  is a good quantifier of flow (element E1) that adheres well to reality (element E2).

### C. Range of the Refined Metric

The most celebrated property of the refined metric is probably its range which is sought in Theorem 1.

**Theorem 1:** Considering both deterministic and probabilistic programs, and all types of an attacker's beliefs, and *avoiding* the imposition of any admissibility restriction on those beliefs, the general range of flow reported by  $Q''$  is:

$$\varrho_{Q''} = [-\eta \cdot \log(1 + b_H(\sigma_H)), \eta \cdot [1 - \log(1 + b_H(\sigma_H))]]$$

where  $\eta$  is the size of a program's secret input in bits.

**Corollary 1:** Notice that  $\log(1 + b_H(\sigma_H)) \leq 1$ . This means that  $Q''_{max} \leq \eta$  and  $Q''_{min} \geq -\eta$ , and makes  $Q''$  size-consistent.

### D. Interpreting the Refined Metric

If we apply formula (11) to the same example given in Section II, we get:

$$Q''(\mathcal{E}, b'_H) = 0.9044 \text{ bits}$$

This time, the flow of 0.9044 bits has brought the attacker from

$$1.5849 \cdot [1 - \log(1 + 0.01)] = 1.5621 \text{ bits}$$

away from reality to

$$1.5849 \cdot [1 - \log(1 + 0.5)] = 0.6577 \text{ bits}$$

away from it. But how much did this flow make the attacker likely to guess correctly? Theorem 2 answers this question, substantiating the validity of Conjecture 1 we made in Section II, and showing that a bit of flow reported by  $Q''$  makes the attacker *more* than twice as likely to guess correctly.

**Theorem 2:** A flow of  $k$  bits reported by  $Q''$  makes the attacker *more* than  $2^k$  as likely to guess correctly. Strictly speaking:

$$Q''(\mathcal{E}, b'_H) = k \Leftrightarrow b'_H(\sigma_H) = 2^{k/\eta} \cdot b_H(\sigma_H) + 2^{k/\eta} - 1 \quad (12)$$

where  $\eta$  is the size of a program's secret input in bits.

### E. Consistency of the Probability Distributions

The bounds of  $Q''$ , given in Theorem 1, ensure *proper* bounds of  $b'_H$ . This can be easily shown by assuming a flow of  $k$  bits and proceeding as follows:

$$\begin{aligned} -\eta \cdot \log(1 + b_H(\sigma_H)) &\leq k \leq \eta \cdot [1 - \log(1 + b_H(\sigma_H))] \\ 2^{\log(\frac{1}{1+b_H(\sigma_H)})} \cdot (1 + b_H(\sigma_H)) - 1 &\leq b'_H(\sigma_H) \\ &\leq 2^{\log(\frac{2}{1+b_H(\sigma_H)})} \cdot (1 + b_H(\sigma_H)) - 1 \\ 0 &\leq b'_H(\sigma_H) \leq 1 \end{aligned}$$

However, this does *not* ensure that an intermediate value of  $Q''$  leads to  $b'_H$  falling outside the range  $[0, 1]$ . To ensure this, we need to show that  $Q''$  is a monotone function. This is done in Lemma 4.

**Lemma 4:**  $Q''$  is a monotonically increasing function, that is:

$$\forall b_1, b_2 : b_1 \leq b_2 \Rightarrow Q''(\mathcal{E}, b_1) \leq Q''(\mathcal{E}, b_2)$$

Thus, the probability distributions dealt with are invariably consistent.

### F. Meaningfulness of the Bounds

We still have to accentuate the meaningfulness of the bounds of  $Q''$  in relation to the attacker's likelihood of a correct guess, or equivalently, to the attacker's certainty about the correct high state. This is done in Theorems 3 and 4.

**Theorem 3:** An *informing* flow equal to the upper bound of  $Q''$  is sufficient to make a fully *uncertain* attacker fully *certain* about the correct high state.

*Corollary 2:* Notice that, in the case of a fully uncertain attacker, we have:

$$\mathcal{Q}_{min}''(\mathcal{E}, b'_H) = -\eta \cdot \log(1 + b_H(\sigma_H)) = -\eta \cdot \log 1 = 0$$

This yields the absolute range  $\varrho_{\mathcal{Q}''} = [0, \eta]$  for  $\mathcal{Q}''$ , and reflects the rationality that a fully uncertain attacker can *only* be informed.

*Theorem 4:* A *misinforming* flow equal to the lower bound of  $\mathcal{Q}''$  is sufficient to make a fully *certain* attacker fully *uncertain* about the correct high state.

A similar corollary to Corollary 2 can be stated to show that a fully certain attacker can *only* be misinformed.

### G. Other Refinements

The discrimination construct, given in formula (9), which we used in our refinement is definitely *not* the only apt construct. Any construct that reduces the discrimination is a likely candidate for the replacement of the Kullback-Leibler construct (given in formula (7)). For instance, consider the following discrimination construct:

$$\mathcal{I}_{Dis}''(\sigma) = \log \frac{1 + b'(\sigma)}{1 + b(\sigma)}$$

This construct clearly cuts down the discrimination. Moreover, it leads to the same refinement that the construct in (9) had led to. This shows that there is a *large* number of possible refinements of the  $\mathcal{Q}$  metric. However, we favored the construct in (9) since the properties of Jensen-Shannon divergence are well-examined in the literature [11].

### V. EXHAUSTIVE SEARCH EFFORT

Assuming a program with a secret input of size  $\eta$  bits, and an informing flow of  $k$  bits from the same program to an attacker. The dynamic upper bound of  $\mathcal{Q}''$ , given in Theorem 1, tells us that  $k \leq \eta$ . Therefore, the space of the exhaustive search [12] that should be carried out in order to reveal the residual part  $\eta - k$  bits of the secret input is  $2^{\eta-k}$ . On the other hand, the dynamic upper bound of  $\mathcal{Q}$ , given in Lemma 1, tells us that  $k > \eta$  is a possible scenario. In scenarios as such, the residual part of the secret input is impossible to determine, and consequently, the exhaustive search space cannot be established, albeit that the secret input might have been partially revealed to the attacker (refer to the example in Section II).

### VI. REMARKS

In addition to the divergence  $K$ , given in formula (10), Lin [11] identified two other divergence measures. The first divergence is denoted as  $J$ , and is given by the formula:

$$J(b \rightarrow b') = \sum_{\sigma \in \mathcal{W}_p} (b'(\sigma) - b(\sigma)) \cdot \log \frac{b'(\sigma)}{b(\sigma)}$$

This divergence is the symmetric form of Kullback-Leibler divergence, given in formula (3), and they both share the same problems; they are unbounded from above and undefined if  $b(\sigma) = 0$  and  $b'(\sigma) \neq 0$  for any  $\sigma \in \mathcal{W}_p$ . It is therefore doubtful that the use of any of these two divergence measures would lead

to size-consistent QIF quantifiers. The second divergence Lin identified is denoted as  $L$ , and is given by the formula:

$$L(b \rightarrow b') = 2S\left(\frac{b+b'}{2}\right) - S(b) - S(b')$$

where  $S$  is Shannon uncertainty functional [5]. This divergence is the symmetric form of the divergence  $K$  we used in our refinement. It has an obvious information-theoretic interpretation in terms of Shannon uncertainty functional which makes it suitable for use in accuracy-based information flow analysis when an attacker's belief about a program's secret input is modeled using advanced representations of uncertainty other than a simple probability distribution over high states. We leave the investigation of this use as future work.

### VII. CONCLUSIONS

We presented a refinement of the QIF metric in [1], [2] that bounds its reported results by a plausible range. Both the original and the refined metric are justified quantifiers of the flow that occurred during a program's execution. However, they differ in their interpretation of one bit of flow. Contrary to the original metric, the results reported by the refined metric are easily associated with the exhaustive search effort needed to uncover a program's secret information (or the residual secret part of that information). We believe that the counter-intuitive flow quantities reported by some QIF quantifiers, that appear in the literature, are due to a flaw in the design of those quantifiers. We further believe that this can be avoided by introducing minor changes into the design of those quantifiers.

### ACKNOWLEDGMENT

The author would like to thank Peter Y. A. Ryan and Marc Pouly for their helpful comments on an early draft of this paper.

### REFERENCES

- [1] M. Clarkson, A. Myers, and F. Schneider, "Belief in information flow," in *Computer Security Foundations, 2005. CSFW-18 2005. 18th IEEE Workshop*, June 2005.
- [2] —, "Quantifying information flow with beliefs," *Journal of Computer Security*, vol. 17, no. 5, 2009.
- [3] G. Smith, "On the foundations of quantitative information flow," in *Foundations of Software Science and Computational Structures*, ser. LNCS. Springer Berlin/Heidelberg, 2009, vol. 5504.
- [4] S. Hamadou, V. Sassone, and C. Palamidessi, "Reconciling belief and vulnerability in information flow," in *Security and Privacy (SP), 2010 IEEE Symposium on*, May 2010.
- [5] J. Y. Halpern, *Reasoning about Uncertainty*. Cambridge, MA, USA: MIT Press, 2003.
- [6] T. M. Cover and J. A. Thomas, *Elements of Information Theory*. Wiley-Interscience, 2006.
- [7] J. Massey, "Guessing and entropy," in *Information Theory, 1994. Proceedings., 1994 IEEE International Symposium on*, Jun-Jul 1994.
- [8] C. D. Manning and H. Schütze, *Foundations of statistical natural language processing*. Cambridge, Mass.: MIT Press, 1999.
- [9] G. J. Klir, *Uncertainty and Information: Foundations of Generalized Information Theory*. Wiley-Interscience, 2005.
- [10] S. Kullback and R. Leibler, "On information and sufficiency," *The Annals of Mathematical Statistics*, vol. 22, no. 1, 1951.
- [11] J. Lin, "Divergence measures based on the Shannon entropy," *Information Theory, IEEE Transactions on*, vol. 37, no. 1, Jan 1991.
- [12] A. J. Menezes, P. C. v. Oorschot, and S. A. Vanstone, *Handbook of applied cryptography*. Boca Raton: CRC, 1997.

## APPENDIX I PROOFS

### A. Proof of Lemma 1

Kullback-Leibler divergence given in formula (3) has the range:

$$0 \leq D(b \rightarrow b') \leq +\infty$$

which means that:

$$\begin{aligned} -\infty &\leq D(b_H \rightarrow \dot{\sigma}_H) - D(b'_H \rightarrow \dot{\sigma}_H) \leq +\infty \\ -\infty &\leq \mathcal{Q} \leq +\infty \end{aligned}$$

It could be safer to bring the reader around by showing the *extreme* cases. The extreme case from above  $\mathcal{Q} = +\infty$  is reached when  $b_H(\sigma_H) = 0$  and  $b'_H(\sigma_H) = 1$ , whereas the converse yields the extreme case from below  $\mathcal{Q} = -\infty$ . An admissibility restriction is suggested in [1] on the attacker's prebelief. This restriction ensures that the prebelief never *deviates* by more than a positive factor from a uniform distribution, and is given by the formula:

$$\min_{\sigma_H}(b_H(\sigma_H)) \geq \epsilon \cdot \frac{1}{|\text{State}_H|}; \epsilon > 0$$

The restriction above more or less excludes the attacker's initial belief that certain states are *impossible*, or in other words, ascribing zero as a prebelief. However, it does *not* impose anything on the attacker's postbelief, which enables us to write:

$$0 < b_H(\sigma_H) \leq 1 \text{ and } 0 \leq b'_H(\sigma_H) \leq 1$$

and consequently:

$$\begin{aligned} -\infty &\leq D(b_H \rightarrow \dot{\sigma}_H) - D(b'_H \rightarrow \dot{\sigma}_H) < +\infty \\ -\infty &\leq \mathcal{Q} < +\infty \end{aligned}$$

Notice how the admissibility restriction is *weak* in that it averts reporting infinite *informing* flow from the metric  $\mathcal{Q}$ , while leaving the rest of the counter-intuitive results unattended (perhaps this explains why the admissibility restriction is given in the original work [1], but not in the expanded one [2]). We have yet to arrive at the general range of  $\mathcal{Q}$ . The last word on this matter relates to the fact that the attacker's postbelief about the correct high state can neither be better than full *certainty* nor worse than full *uncertainty*. The former of these two arguments yields the *dynamic* upper bound of  $\mathcal{Q}$  which corresponds to the maximum *informing* flow:

$$\mathcal{Q}_{max}(\mathcal{E}, b'_H) = -\log b_H(\sigma_H) + \log 1 = -\log b_H(\sigma_H)$$

whereas the latter of the two arguments yields the *absolute* lower bound of  $\mathcal{Q}$  which corresponds to the maximum *misinforming* flow:

$$\mathcal{Q}_{min}(\mathcal{E}, b'_H) = -\log b_H(\sigma_H) + \log 0 = -\infty$$

This gives us the general range of flow reported by  $\mathcal{Q}$ :

$$\varrho_{\mathcal{Q}} = (-\infty, -\log b_H(\sigma_H)]$$

### B. Proof of Lemma 2

The inequality of the arithmetic and geometric means gives us:

$$\frac{b'(\sigma) + b(\sigma)}{2} \geq \sqrt{b'(\sigma) \cdot b(\sigma)}$$

Based on this, we can write:

$$\mathcal{I}'_{Dis}(\sigma) = \log \frac{b'(\sigma)}{\frac{b'(\sigma) + b(\sigma)}{2}} \leq \log \frac{b'(\sigma)}{\sqrt{b'(\sigma) \cdot b(\sigma)}} = \frac{1}{2} \mathcal{I}_{Dis}(\sigma)$$

### C. Proof of Lemma 3

The divergence  $D'$  shown in formula (10) has the range [11]:

$$0 \leq D'(b \rightarrow b') \leq 1$$

which means that:

$$\begin{aligned} -1 &\leq D'(b_H \rightarrow \dot{\sigma}_H) - D'(b'_H \rightarrow \dot{\sigma}_H) \leq 1 \\ \varrho_{\mathcal{Q}'} &= [-1, 1] \end{aligned}$$

### D. Proof of Theorem 1

Borrowing the same two arguments we used in the proof of Lemma 1, we obtain the *dynamic* upper bound of  $\mathcal{Q}''$  which corresponds to the maximum *informing* flow:

$$\mathcal{Q}''_{max}(\mathcal{E}, b'_H) = \eta \cdot [1 - \log(1 + b_H(\sigma_H))]$$

and the *dynamic* lower bound of  $\mathcal{Q}''$  which corresponds to the maximum *misinforming* flow:

$$\mathcal{Q}''_{min}(\mathcal{E}, b'_H) = -\eta \cdot \log(1 + b_H(\sigma_H))$$

This gives us the general range of flow reported by  $\mathcal{Q}$ :

$$\varrho_{\mathcal{Q}''} = [-\eta \cdot \log(1 + b_H(\sigma_H)), \eta \cdot [1 - \log(1 + b_H(\sigma_H))]]$$

### E. Proof of Theorem 2

Assuming a flow of  $k$  bits gives us:

$$\begin{aligned} \mathcal{Q}''(\mathcal{E}, b'_H) &= k \\ \eta \cdot [-\log(1 + b_H(\sigma_H)) + \log(1 + b'_H(\sigma_H))] &= k \\ b'_H(\sigma_H) &= 2^{k/\eta} \cdot b_H(\sigma_H) + 2^{k/\eta} - 1 \end{aligned}$$

### F. Proof of Lemma 4

$$\begin{aligned} b_1 &\leq b_2 \\ -\log(1 + b) + \log(1 + b_1) &\leq -\log(1 + b) + \log(1 + b_2) \\ \mathcal{Q}''(\mathcal{E}, b_1) &\leq \mathcal{Q}''(\mathcal{E}, b_2) \end{aligned}$$

### G. Proof of Theorem 3

A fully uncertain attacker about the correct high state has a zero prebelief. An informing flow equal to the upper bound of  $\mathcal{Q}''$ :

$$\mathcal{Q}''_{max}(\mathcal{E}, b'_H) = \eta \cdot [1 - \log(1 + b_H(\sigma_H))] = \eta \cdot [1 - \log 1] = \eta$$

*evolves* the attacker's knowledge, and transforms her prebelief into the following postbelief:

$$b'_H(\sigma_H) = 2^{k/\eta} \cdot b_H(\sigma_H) + 2^{k/\eta} - 1 = 2^{\eta/\eta} - 1 = 1$$

This postbelief captures the attacker's full certainty about the correct high state.

### H. Proof of Theorem 4

The proof is essentially the same as the proof of Theorem 3, although it starts by a fully certain attacker about the correct high state.